US ATLAS Computing Facilities

Draft 5

1. Introduction

US ATLAS computing facilities are those facilities procured, installed and operated to support US physicists in their participation in the ATLAS project at CERN's LHC. US ATLAS computing facilities are intended to provide the following functions:

- Meet any ***MOU*** defined obligations of US ATLAS to supply direct Information Technology (IT) (computational/network/storage) services to ATLAS.
- Supply IT support for the design, development, construction, calibration and any other activity associated with detector subsystems for which US ATLAS is responsible.
- Supply IT support for US ATLAS obligations to design, develop, test and maintain software, both applications and infrastructure, for ATLAS.
- Supply all IT support beyond the desktop, needed by US ATLAS physicists in analyzing ATLAS data and participating fully in the production of ATLAS physics results.
- Supply such IT resources as required to investigate, prototype, and validate plans for developing optimized facilities, software and infrastructure serving the above functions

The model for US ATLAS facilities is that of a transparent hierarchical distributed grid of computing resources. The components of this Grid are:

- The primary ATLAS Tier 0 computing center at CERN
- A single US ATLAS Tier 1 computing center
- Multiple US ATLAS Tier 2 region computing centers
- Multiple US ATLAS Tier 3 institutional computing facilities
- A dedicated high performance network link between the US ATLAS Tier 1 center and CERN
- Multiple high performance network links connecting together various US ATLAS computing centers and collaborating institutions
- The desktop systems of individual ATLAS collaborator who are members of US institutions

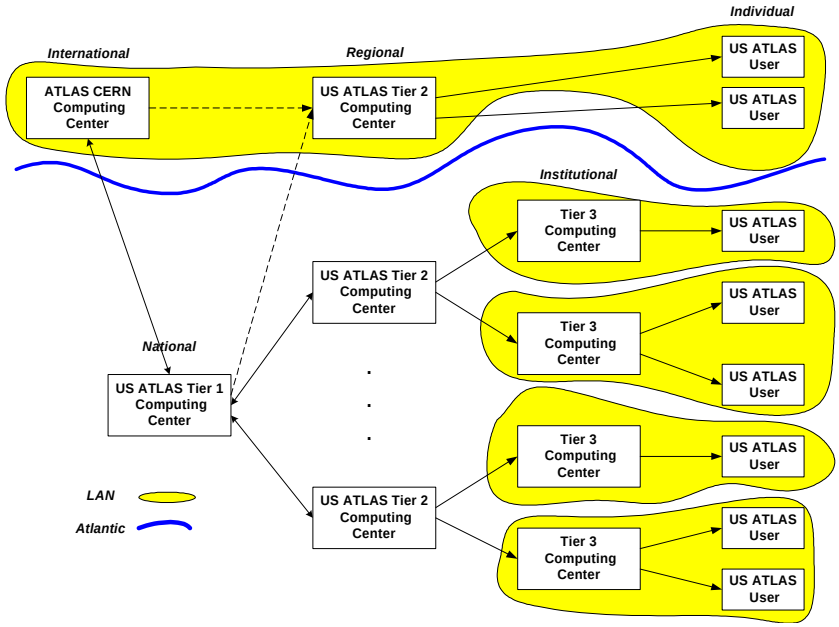


Fig. I Schematic of US ATLAS Computing Facilities Model

In the current model, US ATLAS computing resource will be distributed primarily between Tier 1 and Tier 2 computing centers. There will be a single Tier 1 facility located at Brookhaven National Laboratory, which will be national in scope. This facility will, at least initially, be operated in synergistic conjunction with the RHIC Computing Facility (RCF) currently in operation there. There will be approximately six Tier 2 facilities whose scope will be regional. One of these would likely be located at CERN to support US ATLAS physicists while working there. The sites of other Tier 2 centers remain to be determined but will be selected so as to maximize the number of US ATLAS collaborators who have good access and to maximally leverage local institutional resources and expertise. Tier 3, institutional computing, is expected to be supported primarily from non-ATLAS project sources, with case by case decisions made regarding low levels of project support to facilitate an institutions effective utilization of Tier 1 or 2 resources.

2. Computing Facilities Requirements

The requirements that must be met by US ATLAS computing facilities are highly time dependent. For planning purposes considerable attention has been focused on requirements for the first year of operation, 2005. Clearly as the collider luminosity rises in later years the ATLAS computing requirements will continue to rise but the largest fractional jump in capacity is expected to be for this first year of operation. The computing requirements for 2000 and 2001 can be estimated with reasonable accuracy based on the current people involved and the activities they anticipate for the next couple of years. The requirements for this time frame seem to be quite modest. Year 2002 – 2004 requirements will be associated with the construction, integration and testing of systems and subsystems. This will include optimization of analysis strategies and validation of the computing architecture itself. It is expected that the requirements for this period will be dominated by Mock Data Challenge exercises and by the desire to avoid a facility scale-up of more than a factor of three going into the first year of real operation. ATLAS has not yet defined it Mock Data Challenge plans so some assumptions have been made with will need to be adjust when a formal plan has been established.

The requirements estimate for US ATLAS computing facilities presented here is in derived in part from the requirements established for ATLAS as a whole. The ATLAS assumption is that the main production reconstruction passes on all raw data will be performed at CERN. All other computing functions will be performed in a distributed manner making extensive use of facilities located among the ATLAS collaborating institutions. The general rule of thumb established by ATLAS is that each regional center should supply about 20% of the general ATLAS need. In some cases this distributed computing will be centrally coordinated and done in support of ATLAS in general, for example the production of Monte Carlo data sets. In other cases the computing will be done under local control and will function in support of local physicists, for example the analyses of particular physics channels.

The estimate below was made as follows. The US ATLAS requirement for any particular computing function starts from the ATLAS estimate of that requirement and the estimate made centrally by ATLAS of the fraction of that requirement which is appropriate to a “typical” Tier 1 facility. This initial value is then review from a US ATLAS perspective and on the basis of experience, cost and criticality to US ATLAS participation in the ATLAS physics program this estimate is commonly increased. This revised estimate is then evaluated in the context of the Tier 1 / Tier 2 US ATLAS model. The transition from a monolithic Tier 1 model to a model including Tier 2 centers has to result in some enhancement in overall installed capacity as well as increased collaborator autonomy to justify the increase in technical effort and risk. An increase in the requirement estimate is in some instances made for those functions that are distributed because of the potential for decreased efficiency for distributed resources. The estimating process is discussed below by function for the principle cost contributors of US ATLAS computing facilities.

- Modeling & Simulation: The running of programs modeling various physics processes and the simulated propagation of the resultant particles through the detector will play an important role in understanding the response of the ATLAS detector. In its most complete form the results of such simulations are data sets which are identical in format to actual detector produced raw data. The same analysis chain as is used on actual raw data can then process this simulated data. By so doing one

obtains detailed information about acceptances for signals and backgrounds and is able to most directly confront physics models with measurements. Simulation is a very CPU intensive and ATLAS has estimated the need for about 50x10^3 SPECint95 of installed computing capacity to meet this requirement in 2005. It is assume that US ATLAS will supply 20% of this general ATLAS requirement. In addition, US ATLAS feels it prudent to have the capability to perform a similar amount of simulation in direct support of the analyses of US ATLAS physicists. Thus the simulation requirement for US ATLAS facilities is estimated to be of order 2 x 20% of 50x10^3 SPECint95 or ~20x10^3 SPECint95 of installed capacity. This capacity is something reasonably well suited to relatively lean, in terms of personnel, Tier 2 centers. Therefore in the US ATLAS facilities model a significant fraction of the simulation computing is assigned to Tier 2 centers with a reduction in the simulation capacity which might otherwise have been located at the Tier 1.

- Calibration: While during the construction and testing phase of detectors, for which US ATLAS has responsibility, calibration may be performed by computing resources in the US, it is expected that such calibrations will be done by computing system which are part of the CERN facility once operations begin. In any case this is not expected to be a large load when compared to the simulation, reconstruction or analysis loads. The faction of this load, which actually makes use of U. S. ATLAS resources, is contained within the U.S. ATLAS reconstruction and analysis requirement estimate.

- Reconstruction: The conversion of the digitized values recorded by the detector (raw data) into event summary data (ESD), a description of events in terms of physics variables, is an activity that is expected to occur at least twice for all ATLAS data. It is expected to occur many times more for events of special interest. These reconstruction passes will also produce analysis object data (AOD) which is data optimized for access during final analysis and will contributed to the event tag data (TAG) which enables fast classification of events. Reconstruction is highly CPU intensive and ATLAS has estimated that 70x10^3 SPECint95's of installed computing capacity will be required to satisfy this need in 2005. While the compute resources for the production reconstruction of the Raw data are to be located at CERN, US ATLAS facilities must include the capacity to perform re-reconstruction passes on samples of events of special interest to US ATLAS physicists in doing their analyses. These re-reconstruction passes would usually start from existing ESD but would on occasion go back to the raw data. It is estimate that perhaps 10% of the raw data would be so reprocessed and that the processing load for re-reconstruction from ESD would be about twice this raw data level. Thus the reconstruction requirement for US ATLAS facilities is estimate to be of order 3 x 10% x (70x10^3SPECint95)/2 or ~10x10^3 SPECint95 of installed capacity. A requirement associated with the reprocessing 10% of the raw data is the need to store that amount of raw data in the US. This storage requirement, as will be discussed below, implies that such activities are probably best concentrated at the Tier 1 center.

- Data mining: This is the process performed on ESD or AOD resulting in reduced and/or reorganized ESD or AOD data sets. These sets are better suited in terms of compactness and efficiency of reading for the highly iterative analysis phase of data processing. Data mining is generally a highly I/O intensive process and is not expected to produce a compute load which is large compared to or even clearly distinguishable from the analysis requirement. Data mining is of particular importance for data residing on tertiary storage. The serving of data from both online and tertiary storage is discussed below.

- Data serving (Online storage): This is making the data immediately available for efficient use by processing, compiling, or displaying systems. The data to be served includes documents, code, calibrations, and various levels of event data. It is the serving of the event data that defines the quantitative requirements of the systems. Data serving can be either Local or Wide Area. In the absence of effective differentiated Wide Area Network services, even when there is adequate bandwidth between sites, latency can be a problem. This is particularly true for I/O intensive activities, such as analyses working from database event stores, in which successive reads may be widely separated in storage. It is therefore the goal in US ATLAS computing to have sufficient online storage (disk) Local Area connected to processors, doing any type of computing, so that all of the data

being processed can be cached locally and there is little network latency induced inefficiency. Document and code libraries need to be locally cached but the data volumes associated with this are small and the AFS system handles the caching in an effective way. The capacity defining data to which direct online assess is desired includes all of the TAG, AOD, and ESD for both real and simulated data. Since ATLAS estimates the ESD data volume for 2005 to be of order 100 Tbytes this alone would result in a large cost for disk given that there are 6 Tier 2 as well as a Tier 1 center. US ATLAS physicists are expected to be involved in only a subset of the ATLAS physics analyses and correspondingly in need of frequent access to the associated ESD data. Therefore a model has been used here in which all of the ESD data is available at the US Tier 1 center on tertiary storage but only 50% of it is resident on disk at any time. Calculation based on the above shows that about 100 Tbytes of installed disk would be required to support the complete US ATLAS data serving requirement at the Tier 1 center. In addition significant processing power will be located at the 6 Tier 2 centers, each of which will require significant local disk caches in order to keep its processors *and physicists* operating efficiently. These caches will include additional relatively static subsets of the TAG and AOD, corresponding to the particular analyses concentrated at each site, as well as more dynamic caches of ESD data and other TAG and AOD data. In all, it is estimated that Tier 2 centers may each require as much as 15% of the disk capacity required at the Tier 1 center.

- Nearline storage (Tertiary storage): Tertiary storage will be required to retain copies of all data on disk plus ESD data not on disk and the 10% of raw data for which local re-reconstruction is required each year. This results in a data volume, which is of order 300 Tbytes in 2005. In addition to the automated media library, which stores and manipulates the recording media, there are drives, which do the actual reading and writing. The number and performance of these drives determine the access bandwidth to the data store. An estimate of the bandwidth requirement for access to these data is obtained by estimating that one would want to be able to make passes through the ESD data not on disk at least once a week so 50 Tbytes/ 7 days or ~100 Mbytes/sec. Assuming that the aggregate throughput for all other activities is comparable, the installed tertiary I/O bandwidth must be of order 200 Mbytes/sec. The US ATLAS model is that tertiary storage with the associated high personnel requirements of hierarchical storage system hardware and software management will be limited to the Tier 1 center.

- Analysis: This is a highly iterative process to extract physics significance from a data set by the adjustment of selection criteria and display variables. This activity is most commonly interactive in nature with the physicist actually sitting and waiting for the result. For this reason the requirement is defined by peak demands where the peaks occurring during the workday hours of geographical locations where large concentrations of ATLAS collaborators are found. ATLAS has estimated a requirement for 130 kSPECint95 of capacity for such analysis. The US ATLAS community is distributed across a 3-hour range in time zones and so for 6-8 hours each day there is likely to be heavily overlapped use of these facilities by almost all active US ATLAS collaborators and therefore the demand very peaked. The rule of thumb indicates that US ATLAS resources should include 20% of the total analysis capacity. However, because of this strong peak in demand and since this is the compute capacity that directly supports US ATLAS physicists in the extraction of results for the physics channels they are studying, the requirement for analysis has been raised to twice the guideline or 40%. This produces an estimate for analysis of 2 x 20% x 130x10^3 SPECint95 = ~50x10^3 SPECint95 of installed capacity. This function is well suited to Tier 2 centers and is expected to be distributed there with a substantial descoping of this activity at the Tier 1 center.

- Data import & export: ATLAS has estimated that the required data transfer rate between CERN and a Tier 1 Regional Center could be satisfied by a dedicated OC3 connection. The required connectivity for bulk data transfer between multiple Tier 2 centers and the US Tier 1 is expected to be at a comparable bandwidth. Since OC3 connectivity is already becoming common within the US and is expected to be available to CERN soon, the use of physical media to export and import data is unlikely to be necessary and so has not been included in this model.

- Desktops: Functionally this includes Email, web browsing, document preparation, image display, etc. which are currently the purview of the individual physicists desk top workstation or PC. This activity is largely independent of the scale of the project on which the physicist works. While it involves

Information Technology it is more of an infrastructure appliance such as the telephone and so is assumed to exist but is not considered a part of the US ATLAS computing requirement proper.

- Network: Connectivity between CERN and the Tier 1 center was estimated to require at least an OC3 connection. While OC3 allows for the time averaged bulk transfer of data from source sites to other required centers, US ATLAS feels that it is marginal in terms of dealing with peak demand situations. It will further put an unacceptable premium on managing data flow in the dynamic world of replicated data bases being simultaneously manipulate at multiple sites. US ATLAS therefore estimates that by the time ATLAS operations begin, OC12 connectivity for US ATLAS to CERN well be required. By the same logic, connectivity between the Tier 1 center and Tier 2 centers will also require OC12.

- Infrastructure Software: There are a large number of functionality requirements, which impact software more than hardware. Such software requirements can result in significant license and maintenance cost as well as staffing cost related to development, integration, and operations efforts. A few of the more important items included in these requirements are: data management tools such as an OODBMS, wide area network transparency, software development tools and utilities, code and document repository management utilities, and the cyber security of user data and facility resources.

3. Technology and Technology Projections

Virtually all of the cost driving components of the US ATLAS facilities plan are evolving dramatically as a function of time. This means that any projection of cost for a specified level of performance at a time five years in the future is going to have large uncertainties associated with it. The following describes the technology choices and costing assumptions of these major facility cost-driving components.

- Compute intensive (CPU): The requirement for large quantities of inexpensive compute cycles has been most effectively met over the course of the last 20 years by farms of semi-commodity (workstations) and more recently commodity (personnel computer) computers. The decrease in the cost of CPU cycles from these commodity systems has continued at the rate of at least a factor of two every 18 months for more than 15 years and shows no sign of stopping. It is therefore clear that in the near term such farms will play a part in ATLAS computing and it is seems likely that this will remain to be the case into initial LHC running. Computing of this type is extremely well suited to the simulation and reconstruction component of the ATLAS computing requirement. Farms of this type based on Intel processors running the Linux operating system are currently in wide use and are the costing basis for this type of computing.

- Online storage (Disk): While various other technologies have shown promise, magnetic disk with a decrease in cost of approximately a factor of 2 every 18 months has remained the primary form of Online storage for several decades. There are strong indications that it will retain its rate of improvement in cost effectiveness and remain the principal online storage medium for the near term. It replacement by another technology would be a result of the technology improving even more rapidly so estimating costs on the basis of magnetic disk technology is a conservative approach. Fibre Channel connected RAID disk is the costing base here for this type of storage.

- Nearline (tertiary) storage (Robotic tape): While the rate of improvement in cost effectiveness for tape storage has not been as dramatic as that of CPU and disk, its improvement has been significant during the last few years. A number of tape technology advances in capacity and bandwidth currently predicted offer the possibility that over the next couple years it price performance improvements may approach those of disk and CPU, factors of two in 18-24 months. The management of the large volumes of data stored in an automated tape storage system requires the use of complex software for both bookkeeping and operations. This software usually couples the automated tape library, the tape drives, and a front-end disk cache into a unified hierarchy of storage. The High Performance Storage System (HPSS), a Hierarchical Storage Management system distributed by IBM, is currently perhaps the only highly scalable HSM commercial used in particle physics and so is a reasonable default for the US ATLAS model. It allows for reasonably accurate estimation costs and staffing requirements.

- Network: The introduction of first fast Ethernet and then Gigabit Ethernet have resulted in dramatic improvement in both performance and price/performance over the course of the last few years for local area networking. The network requirements within US ATLAS facilities seem to be well within the capacity of such technology. In the case of wide area networking the situation is less clear. The cost of wide area networking is currently substantial, especially between the US and Europe. Simple extrapolations of existing technology are used to estimate the availability and cost of these capabilities. At this time these numbers are highly uncertain and should be considered as placeholders for an important and possibly quite expensive component of the US ATLAS facilities.

- Servers: Medium to large scale SMP's with their high speed back planes and their capability of supporting large numbers of peripherals and network connections are available from SUN, IBM, and SGI. All make SMP's commonly used in HEP as server machines. Such severs are also frequently used to support I/O intensive activities such as data mining or some types of analysis. While such system are experiencing improvement in price/performance at approximately the same rate as commodity computers, their price is offset higher by a factor of 5 to 10 so it is important that they not be used for simple CPU intensive activities. These systems fill the large-scale server niche in the US ATLAS facilities model. One of the major server functions of these SMP's, that of serving RAID disk, may be usurped by dedicated serving engines, so called network appliances, over the course of the next few years. However since this will only occur if the new technology is more cost effective than the continued use of SMP servers, the conservative costing based on SMP's is used here.

4. Architecture

As discussed briefly above the model for US ATLAS facilities is that of a transparent hierarchical distributed grid of computing resources. The simplest form of the model is one in which the only significant concentration of resources that exist outside of CERN would be at the Tier 1 center. This degenerate form of the model might occur in a limited funding scenario. It would quite efficiently address the basic requirements but would make less effective use of desktop or other institutional or base program supplied resources and would, with the associated need for formal centralize resource allocation, reduce the autonomy of individual technical and scientific work groups.

At the other end of the model spectrum is one where comparable capacities exist at each level of the hierarchy. While the utilization of such more distributed resources will be less efficient, it is expected to be offset by the ability to utilize institutional, desktop and other base program supported equipment more effectively. Effective utilization of such a full function hierarchy requires more technical support and in particular the use of analytic and predictive tools such as those being developed in MONARC, see Appendix A, and the successful execution of advance GRID computing projects such as the Particle Physics Data Grid and Apogee. The result would be a computing environment with more technical capacity local to individual institutions, thus putting more functionality into hands of individual physicists and granting significantly more autonomy to technical and scientific work groups.

The US ATLAS facilities plan focuses on a computing hierarchy in which there are significant resources at all levels. However attention has been paid to assuring that the overall design will work efficiently in the absence of certain intermediate levels. This would require quantitative changes in the composition of the remaining levels of the hierarchy but would not require dramatic changes in architecture or changes in the fundamental capabilities at any level.

Although it is not immediately obvious, for a computing facility offering a full range of services, the primary cost associated with supplying computing service is almost always the cost of the personnel required to operate the facility rather than the cost of the computing hardware in it. A quick evaluation of the US ATLAS case leads to this same conclusion. Therefore one of the guiding principles in developing the model for US ATLAS computing facilities has been an effort to limit the scale of personnel requirements. One way to address the personnel problem in a multi-center facilities plan is to make only one of those centers a full function center in need of a full complement of expertise. The second is to enforce a high level of uniformity across the various centers so that a solution developed for one is

immediately applicable to the others. Both of these approaches are part of the plan for a multi-center U.S. ATLAS computing facilities architecture.

The Tier 1 regional center for US ATLAS located at Brookhaven National Laboratory will contain all elements of US ATLAS facilities and will supply full support services including mirrors of the ATLAS code and documentation repositories at CERN. The Tier 2 centers will focus on simulation and analysis compute capabilities with sufficient local disk cache to efficiently support these activities. They will specifically not include tertiary storage systems, depending on high performance network connections and GRID computing infrastructure to make us of such capabilities at the Tier 1 center.

5. Facility Components

There are three basic cost driving components to US ATLAS facilities. The first is the Tier 1 center at BNL. The second is the set of Tier 2 facilities. The third is the network. The Tier 1 facility is already coming into operation and will serve as the initial test bed for US ATLAS facilities, their configuration and how they interact with users. The Tier 2 facilities will be phased in. On the assumption that it takes two years to achieve fully operational status, the first, serving as the Tier 2 testbed will be established in 2001, the second in 2002, and the remaining 4 in 2003 so that all will be fully operational going into 2005. The second Tier 2 center, established in 2002 would probably be the one at CERN. The 2005 projected capacities of the Tier 1 Center and a typical Tier 2 center are indicated in the tables below.

Tier 1 Center in Context of Tier 2's

	CERN Disk	CERN Tape	US Fractions		US ATLAS Disk	US ATLAS Tape
Annual Data Storage Estimate						
Raw - Tbytes		1000	10%			100
ESD - Tbytes	100	100	100%		50	100
AOD+ETD - Tbytes	11	11	100%		11	11
Simulated - Tbytes	11	111			6	21
10% Tape Staging & Caching - Tbytes					23	
Total Storage for 2005 Data - Tbytes	122	1222			90	232
Legacy Sim & T-beam Data ('99-'04) - Tbytes					10	75
Total Installed Capacity in 2005 - Tbytes					100	307

	GLOBAL ATLAS	US Fraction	US ATLAS
Installed CPU Estimate			
Simulation - kSPECint95	50	20%	10
Reconst. - kSPECint95	70		
US Reconst. - kSPECint95		10%	7
Analysis - kSPECint95	130	20%	26
Total CPU - kSPECint95	250		43

	Comment	Band-width
CERN connectivity - Mbits/sec	OC12	622
User connectivity - Mbits/sec	50 Users	50
US Tier 2 connectity - Mbits/sec	5 OC12 SITES	3110

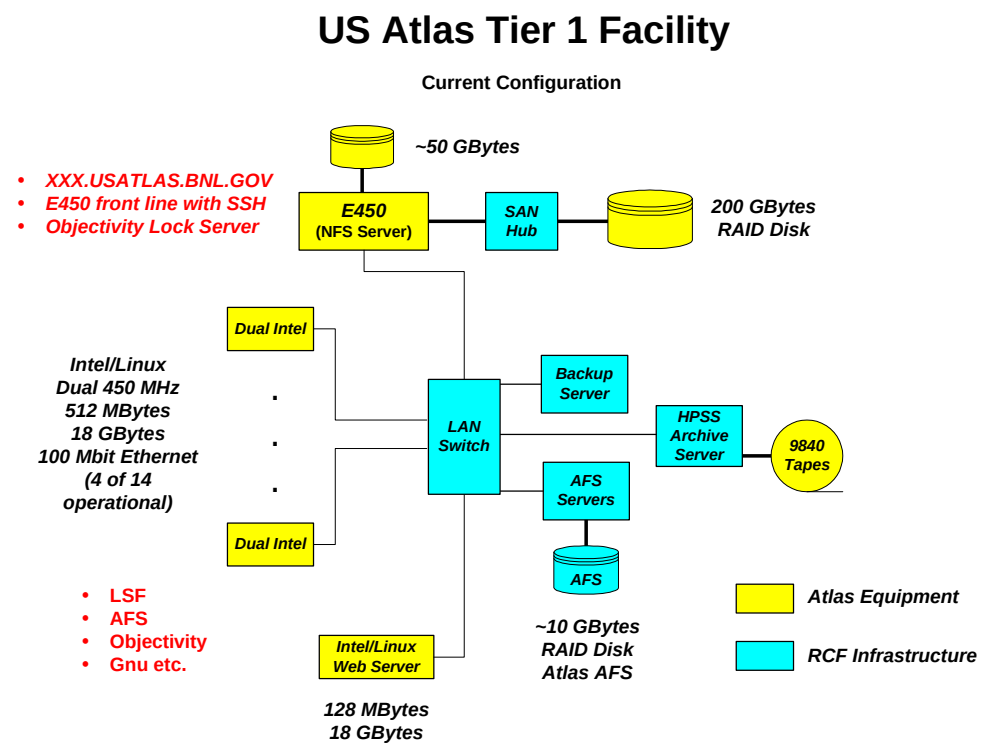
Table 1

Typical 1 of mutiple Tier 2 Centers			
	Comment		
CPU			
Simulation - kSPECint95	20%	of above US	4
Analysis - kSPECint95	20%	of above US	10
Disk			
AOD-EOT Subsets - Tbytes	50%	of Total	6
Network-Tape Stage & Cache - Tbytes	3%	of US Tape	9
Network			
Tier 1 connectivity - Mbits/sec	Dedicated OC12		622
User connectivity - Mbits/sec	20%	of US users	20

Table 2

The Tier 1 facility at BNL is currently operational with several of the basic components but with very limited capacity in each. The facility currently consists of a Sun server, Fibre Channel connected RAID disk, a small farm of dual processor Intel systems running Linux, and a web server. The facility shares physical space and external network connectivity with the RHIC Computing Facility (RCF). The US ATLAS facility also has shared access to a low priority HPSS class of service for the archiving of data and shares use of an AFS server at the level of about 10 GBytes of storage. A schematic is shown below.

Fig. 2



There is currently one FTE's of system administration effort being directed to its support, which is covered by BNL LDRD funding. The current installed capacities are summarized as follows:

Table 3
Current U.S. ATLAS Tier 1 Capacities

Compute	28 CPU's	500 SPECint95
Disk	Fibre Channel	250 Gbytes
Sun Server / NIC	2 CPU's	100 Mbit/sec

6. Capabilities and Cost

The following table describes proposed US ATLAS computing capacities as a function of US Fiscal year for CPU, Online (Disk) Storage, and Nearline (Tertiary) Storage. The table shows the break down between the Tier 1 center and Tier 2 centers. The less significant digits in the estimated quantities are a result of the computational process by which they were obtained and should not be regarded as significant.

US ATLAS Facilities Integrated Capacities

	FY 1999	FY 2000	FY 2001	FY 2002	FY 2003	FY 2004	FY 2005	FY 2006
Operational Tier 2 Facilities	-	-	1	2	6	6	6	6
CPU - kSPECint95								
Tier 1	0.2	0.5	1.0	3	6	17	50	83
Tier 2	-	-	1.0	3	12	30	89	154
Total CPU	**0.2**	**0.5**	**2.0**	**6**	**18**	**47**	**140**	**237**
Disk - TB								
Tier 1	0.2	0.2	2.0	5	13	34	100	169
Tier 2	-	-	0.6	3	12	28	89	147
Total Disk	**0.2**	**0.2**	**2.6**	**8**	**25**	**62**	**189**	**316**
Robotic Tape Storage (Tier 1) - TB								
Total Robotic Tape	**1**	**5**	**11**	**20**	**34**	**101**	**304**	**607**

Table 4

The personnel requirements for the Tier 1 and Tier 2 facilities in 2005, presumably near steady state operating mode, and how they are imagined to be distributed according to function are indicated in the following table. While the detail is again not to be taken very seriously, it is felt that integrating over the details will produce estimates that are reasonably accurate. As has been discussed elsewhere, the goal with the Tier 2 centers has been to limit the required staffing and to try to leverage existing facilities and expertise at institutions in site selection. For this reason, while it is felt that not fewer than 3 FTE's are required to operate such a facility, at least one of those three is expected to be supported out of the local base program or by some other institutional contribution. Therefore only 2 FTE's per Tier 2 site are expected to contribute to the final project cost.

Table 5

US ATLAS Facilities FY 2005 Staffing Estimate Details

Function	Typical Tier 2 (FTE's)	Multi-center Tier 1 (FTE's)	Total 6 x Tier 2 (FTE's)
GRID / Distributed System	0.5	4	3.0
Computing Environment	0.5	5	3.0
Simulation/Reconstruction Systems	0.4	2	2.4
Analysis Systems	0.5	2.5	3.0
Data Storing & Serving	0.4	5	2.4
Network	0.2	2	1.2
Measure & Monitor Performance	0.2	2.5	1.2
Management	0.3	2	1.8
Total	**3.0**	**25**	**18**
	Assume 1 FTE is contributed from base		Total of 6 FTE's are contributed from base

The funded staffing profile for each type of center and integrate across all centers is shown by year in the following table.

Table 6

US ATLAS Facilities Staffing (FTE's)

	FY '99	FY '00	FY '01	FY '02	FY '03	FY '04	FY '05	FY '06
Tier 1								
Tier 1 Total	**1**	**4**	**8**	**11**	**16**	**19**	**25**	**25**
Tier 2								
Initial Year Center	-	-	1	2	2	2	2	2
Second Year Center	-	-	-	1	2	2	2	2
4 Final Year Centers	-	-	-	-	4	8	8	8
Tier 2 Total	**-**	**-**	**1**	**3**	**8**	**12**	**12**	**12**
US ATLAS Facilities Total	**1**	**4**	**9**	**14**	**24**	**31**	**37**	**37**

Network costs will be a significant contribution unless they are funded out of some other source. Since adequate network capacity is critical to the US ATLAS Facilities plan an estimate of the cost of the required network capacity as summarized below has been included. This estimate is highly uncertain.

Table 7

Network Cost Estimate

	1999	2000	2001	2002	2003	2004	2005	2006
Tier 1 to CERN Link					T3	OC3	OC12	OC12
Annual CERN Link Cost	**0**	**0**	**0**	**0**	**200**	**300**	**400**	**300**
Number of Tier 2 to Tier 1 OC3 Links	0	0	1	2	5	4	0	0
Number of Tier 2 to Tier 1 OC12 Links	0	0	0	0	0	1	5	5
Estimate annual cost of a domestic OC3	250	200	160	128	102	82	66	52
Estimate annual cost of a domestic OC12	500	400	320	256	205	164	131	105
Total Domestic WAN Cost	**0**	**0**	**160**	**256**	**512**	**492**	**655**	**524**
Total WAN Cost	**0**	**0**	**160**	**256**	**712**	**792**	**1055**	**824**

The following table summarizes the total US ATLAS facilities costs as a function of U.S. Fiscal Year. It should be noted that certain items which are usually separated out from equipment, such as power and maintenance costs are here included in the Equipment, etc. line.

Table 8

US ATLAS Facilities Annual Costs (FY2000 k$)

	FY '99	FY '00	FY '01	FY '02	FY '03	FY '04	FY '05	FY '06
Tier 1								
Equipment, etc.	110	220	560	590	910	1,650	2,450	2,130
Personnel	30	560	1,120	1,540	2,230	2,650	3,490	3,490
Tier 1 Total	150	780	1,670	2,130	3,150	4,310	5,950	5,620
Tier 2								
Equipment, etc.	-	-	190	380	1,150	1,380	2,580	2,110
Personnel	-	-	140	420	1,120	1,680	1,680	1,680
Tier 2 Total	-	-	330	800	2,270	3,060	4,260	3,790
Network								
Network Total	-	-	160	260	710	790	1,060	820
US ATLAS Facilities Total	**150**	**800**	**2,200**	**3,200**	**6,100**	**8,200**	**11,300**	**10,200**

7. Milestones and Deliverables

The deliverables associated with the facilities component of the US ATLAS computing project are most simply defined as the capacities projected for each year in Table 4 above. To these one must add appropriate network connectivity and properly installed and maintained infrastructure software as well as the required tools, utilities and libraries. There are in addition qualitative aspects to these requirements, which have at this point not been fully defined but need to be reflected in the ultimate list of deliverables. Certainly the notion of transparency of use of facilities in this distributed model is an important qualitative aspect of this type. Others will include ease of use and general uniformity in look and feel. A different class of deliverable is represented by the required support of US ATLAS physicist in their use of the facilities. Since requirements will change over the course of the next five years, it is necessary that the deliverables be adjusted to meet those revised requirements.

Milestones associated with the facilities component of the computing project are also complex and not yet fully defined. The milestones will include the making of technical and managerial decisions, demonstration of developed and procured technical capabilities, and the establishing of levels of capacity. Table 9 below contains a strawman list of potential milestones.

Table 9

Draft Facilities Milestone List

Milestone Description	Date
Selection of 1st Tier 2 site	01-Oct-00
Procure Automate Tape Library (ALT)	01-Jun-01
Demo Tier 2 transparent use of Tier 1 HSM	01-Jan-02
Establish dedicated Tier 1 / CERN link	01-Jan-03
Select remaining (4) Tier 2 sites	01-Jan-03
Mock Data Challenge I (10% turn-on capcity)	01-May-03
Final commit to HSM	01-Oct-03
Mock Data Challenge II (33% turn-on capacity)	01-Jun-04
Achieve turn-on capacities	01-Jan-05

Appendix A

MONARC

The MONARC project is the means by which the experiments have banded together to meet the technical challenges posed by the storage, access and computing requirements of LHC data analysis. The baseline resource requirements for the facilities and components of the networked hierarchy of centres, and the means and ways of working by which the experiments may best use these facilities to meet their data-processing and physics-analysis needs, are the focus of study by MONARC. Tufts University (K.Sliwa) on the ATLAS side and Caltech (H. Newman) on the CMS side led the joint effort of creating this common CERN project.

The scale, complexity and worldwide geographical spread of the LHC computing anddata analysis problems are unprecedented in scientific research. Each LHC experiment foresees a recorded raw data rate of 1 PetaByte/year (or 100 MBytes/sec during running) at the start of LHC operation. This rate of data to storage follows online filtering by a factor of several hundred thousand, and online processing and data compaction, so that the information content of the LHC data stores will far exceed that of the largest PetaByte-scale digital libraries foreseen for the next 10-15 years. As the LHC program progresses, it is expected that the combined raw and processed data of the experiments will approach 100 PetaBytes by approximately 2010. The complexity of processing and accessing this data is increased substantially by the size and global span of each of the major experiments, combined with the limited wide area network bandwidths that are likely to be available by the start of LHC data taking.

The general concept developed by the two largest experiments, CMS and ATLAS, is a hierarchy of distributed Regional Centres working in close coordination with the main centre at CERN. The regional centre concept is deemed to best satisfy the multifaceted balance needed between

- proximity of the data to centralised compute and data handling resources,
- proximity to the end-users for frequently accessed data,
- efficient use of limited network bandwidth,
- appropriate exploitation of regional and local computing and data handling resources,
- effective involvement of scientists in each country and each world region in the data analysis and the realisation of the experimental physics discoveries.

The use of regional centres is well matched to the worldwide-distributed structure of the collaboration, and will facilitate access to the data through the use of national and regional networks of greater capacity than may be available on intercontinental links.

The primary goals of MONARC are to:

- determine which classes of models, and modes of distributed analysis, are feasible according to the network capacity and data-handling resources available at the collaborating sites
- specify the main parameters that characterise these classes of models
- produce example baseline models which fall into the "feasible" category
- deliver a set of tools for simulating candidate computing models of the experiments
- formulate a set of common guidelines to allow the experiments to formulate their final Models
- formulate a set of guidelines for Regional Centre architecture and functionality, as well as the interactions among the Centres

In order to achieve these goals MONARC has organised itself into four working groups, and is led by a Steering Group responsible for directing the project and coordinating the Working Group

activities. Members of the Steering Group are given below:

Steering Group Member	Principal Activity
Harvey Newman (Caltech)	Spokesperson
Laura Perini (INFN Milano)	Project Leader
Krzysztof Sliwa (Tufts)	Simulation and Modelling WG Leader
Joel Butler (Fermilab)	Site and Network Architectures WG Leader
Paolo Capiluppi (INFN Bologna)	Analysis Process Design WG Leader
Lamberto Luminari (INFN Roma)	Testbeds WG Leader
Les Robertson (CERN IT)	CERN Centre Representative
David O. Williams (CERN IT)	Network Evolution and Costs
Frank Harris (Oxford/CERN)	LHCb Representative
Luciano Barone (INFN Roma)	Distributed Regional Centres
Jamie Shiers (CERN IT)	RD45 Contact
Denis Linglin (CCIN2P3 Lyon)	France RC Representative
John Gordon (RAL)	United Kingdom RC Representative
Youhei Morita (KEK)	Objectivity WAN (KEK)

A Regional Centres Committee has been formed, composed of representatives of actual and potential regional centres; which acts as an extended MONARC Steering Group.
The MONARC Project has accomplished its primary goals of identifying baseline Computing Models that could provide viable (and cost-effective) solutions to meet the data analysis needs of the LHC experiments, providing a simulation toolset that will enable further Model studies, and providing guidelines for the configuration and services of Regional Centres. The criteria governing the MONARC work are:

- the network bandwidth, computing and data handling resources likely to be available at the start of and during LHC running,
- the computing power and data transport speeds needed for an effective data analysis,
- the features and performance of the distributed database system and
- an overall strategy for data processing, distribution and analysis that meets the needs while using the resources efficiently, with acceptable turnaround times.

The main deliverable from the project is a set of example "baseline" Models. The project aims at helping to define regional centre architectures and functionality, the physics analysis process for the LHC experiments, and guidelines for retaining feasibility over the course of running. The results will be made available in time for the LHC Computing Progress Reports, and could be refined for use in the Experiments' Computing Technical Design Reports by 2002.
The approach taken in the Project is to develop and execute discrete event simulations of the various candidate distributed computing systems. The granularity of the simulations is adjusted according to the detail required from the results. The models are iteratively tuned in the light of experience. The model building procedure, which is now underway, relies on simulations of the diverse tasks that are part of the spectrum of computing in HEP. A simulation and modelling tool kit has been developed and validated against the measurements performed with test-beds. Also, the simulation results were shown to exactly reproduce analytical predictions based on queuing theory.
As scheduled in the PEP, the **MONARC Simulation WG** has developed a flexible and extensible set of common modelling and simulation tools. These tools are based on Java2, which allows the process-based simulation system to be modular, easily extensible, efficient (through the use of multi-threading) and compatible with most computing platforms. The system is

implemented with a powerful and intuitive Web-based graphical user interface that will enable MONARC, and later the LHC experiments themselves, to realistically evaluate and optimise their physics analysis procedures. Iosif Legrand was the primary developer of the simulation tool. Alex Nazarenko of Tufts University has joined him in the Summer of 1999, and helped develop the improved gui and the built-in set of statistical data analysis tools. The physics activities steps which can be simulated at present are:

- **Reconstruction**. The process has to be performed at the Off-line Farm at CERN for all the WGs. This in fact means the filling process of the Objects in the Object Database. Possible re-reconstructions are one of the parameters of the Model, including their possible location (either at CERN or partially at Regional Centres). The so called ESD are produced during these processes. Data produced are of the order of 100 TBytes/year and they reside also in the Regional Centres for the part needed by the "regional" activities. Disk storage media are foreseen for this type of (output) data sample. Tapes may be also needed, depending on cost and technology evolution.
- **Selection**. The data-sample is selected and reduced in size and number of events, eventually in two subsequent Passes triggered by individual Groups, in order to provide the database information relevant for the analysis. This is the more relevant and delicate process, producing the so called AOD. Data produced are evaluated for different selections. The results are strongly dependent on the number of "passes" and designed activities, ranging from final 2TB/year to 0.2 TB/year for the whole experiment. Disk storage at the Regional Centres should be the choice for these data samples.
- **Analysis**. The group-produced data sample is inspected by individual components so as to obtain physics results. Simulated data will also be used during this process. Data samples will certainly be stored on disks and the jobs will run at the Regional Centres. The possibility of undertaking part or all of this activity on Institute resources (Desktops) is under evaluation.
- **Simulation.** The model includes the distributed production of Monte Carlo event simulation, and the reconstruction. The current practise in HEP experiments of distributing and coordinating simulation is well established: this fact led us to retain distributed simulation in the LHC computing model. Group simulations may use dedicated (Tier2 Regional Centres) and/or distributed resources available to the Collaboration.

Tufts University Group (Nazarenko and Sliwa) has developed a complete model of all physics activities foreseen by the LHC experiments, and then proceeded to simulate the two possible computer system architectures – fully centralized (one center at CERN) and a distributed system with multiple Tier-1 and possibly Tier-2 centers. With the completion of this task in December 1999, the primary MONARC goals for its first two stages have been fulfilled. In the last LCB meeting it was agreed that a detailed user's manual would be the preferred way to summarize the results. Tufts University group is leading this project.

The **Site and Networks Architectures WG** has studied the computing, data handling and I/O requirements for the CERN centre and the main "Tier1" Regional Centres, as well as the functional characteristics and wide range of services required at a Tier1 Centre. A comparison of the LHC needs with those of currently running (or recently completed) major experiments has shown that the LHC requirements are on a new scale, such that worldwide coordination to meet the overall resource needs will be required. Valuable lessons have been learned from a study of early estimates of computing needs during the years leading up to the "LEP era". A study of the requirements and modes of operation for the data analysis of major experiments just coming (or soon to come) into operation has been started by this group. The group is also beginning to develop conceptual designs and drawings for candidate site architectures, in cooperation with the MONARC Regional and CERN Centre representatives.

The **Analysis Process Design WG** has studied a range of initial models of the analysis process. This has provided valuable input both to the Architectures and Simulation WG's. As the models and simulations being conducted became more complex, close discussions and joint meetings of the Analysis Process and Simulation WG's began, and will continue. In the future, this group will be responsible for determining some of the key parameter sets (such as priority-profiles and breakpoints for re-computation versus data transport decisions) that will govern some of the large scale behaviour of the overall distributed system.
The **Testbeds WG** has defined the scope and a common (minimum) configuration for the testbeds with which key parameters in the Computing Models are being studied. The recommended test environment including support for C++, Java, and Objectivity Version 5 has been deployed on Sun Solaris as well as Windows NT and Linux systems. A variety of tests with 4 sets of applications from ATLAS and CMS (including the GIOD project) have begun. These studies have being used to validate the simulation toolset as well as extracting key information on Objectivity performance.
Distributed databases are a crucial aspect of these studies. Members of MONARC also lead or participate in the RD45 and GIOD projects which have developed considerable expertise in the field of Object Database Management Systems (ODBMS). The understanding and simulation of these systems by MONARC have benefited from the cooperation with these projects.

STATUS OF MONARC SIMULATION TOOL

Requirements

The development of a powerful and flexible simulation and modelling framework for the distributed computing systems was the most important task in the first stage of the project. Some requirements for the framework are listed below:

- allow the study of candidate reconstruction and analysis process architectures for the LHC experiments,
- perform reliable modelling of large computing facilities and the networks connecting them,
- carry out simulations of complex models as fast as possible without jeopardising the correctness of the results
- the model implemented in the framework's simulation program should be equivalent in behaviour to the real system in all important aspects
- understanding of the real system's components in terms of the simulation model should be straightforward.

The distributed nature of the reconstruction and analysis processes for the LHC experiments required the framework's simulation program capable of describing complex patterns of data analysis programs running in a distributed computing system. It was recognised from the very beginning that a process-oriented approach for discrete event simulation is well suited to describe a large number of programs running concurrently, all competing for limited resources (data, CPU, memory, network bandwidth etc.).

Implementing the Data Model

As envisaged in the Computing Proposals of the LHC experiments, all data is organised in objects and managed within the framework of an object database. In our case we consider specifically an Objectivity/DB federated database, which allows to distribute sets of objects onto different media, geographically and physically, media (tape, disk...) and data servers (Objectivity AMS servers), while maintaining a coherent and logically uniform view of the entire distributed database. Objects can contain pointers to each other (associations) which enable navigation across the entire database. The data model implemented in the simulation consists of 4 functionally different groups of objects:

- RAW data; about 1MB/event, most likely to be stored on tape only at CERN
- ESD data (Event Summary Data) - objects with reconstructed information; about 0.1 MB/event
- AOD data (Analysis Object Data) - a subset of ESD (possibly non-overlapping, connected via AOD->ESD associations); about 0.01 MB/event
- TAG - a small set of essential information describing a physics event (jet and lepton multiplicity, trigger masks, values of the transverse energy of the most energetic jets and leptons...) which allows initial selections of which AOD data to process

Data of these four different types are organised in unique containers (files). The simulation has a software equivalent of a real Objectivity/DB database catalogue, which allows each job to identify which containers are needed for processing the data requested by that JOB. The locking mechanism has been implemented on the container level, as in Objectivity federated databases. Different types of operation on the data are modelled by different JOBS; for example RAW->ESD, ESD->AOD and AOD->TAG processing involves different input and output data, and different processing time. For example, if the initial FARM configuration has all data on TAPE, if RAW->ESD jobs are submitted to the queues, they invoke the TAPE->DISK copy process.

JAVA2–based MONARC simulation tool

The scheme, developed using Java2 tools, provides for an efficient way to handle a very large number of objects and automatic storage management, allows one to emulate different clustering schemes of the data for different types of data access patterns as well as to simulate the order of access following the associations between the data objects, even if the objects reside in databases in different AMS servers.The NETWORK model has been modified as well. It is, at present, an "interrupt" driven simulation. For each new message an interrupt is created, which triggers a re-calculation of the transfer speed and the estimated time to complete a transfer for all the active objects. Such a scheme provides an efficient and realistic way to describe (simulate) concurrent transfers using very different object sizes and protocols. Logically, there is no difference in the way LANs and WANs are simulated. A multi-tasking processing model for shared resources (CPU, Memory, I/O channels) has been implemented. It provides an efficient mechanism to simulate multitasking and I/O sharing. It offers a simple mechanism to apply different load balancing schemes. With the new program it is now possible to build a wide range of computing models; from the very centralised (in which the reconstruction and most analyses are performed at CERN) to the distributed systems, with an almost arbitrary level of complication (CERN and multiple regional centres, each with different hardware configuration and possibly different sets of data replicated). A much improved GUI, enhanced graphical functions and built-in tools to analyse results of the simulations are also provided. In table below all parameters currently in use by the MONARC simulation tool are listed.

Table 1. Parameters used by the MONARC simulation tool.

federated database and data model parameters (global)	Regional centre configuration parameters (local)
Database page size	Number of AMS_servers
TAG object size/event	AMS link speed
AOD object size/event	AMS disk size
ESD object size/event	Number of processing nodes
RAW object size/event	CPU/node
Processing time RAW->ESD	Memory/node
Processing time ESD->AOD	Node link speed
Processing time AOD->TAG	Mass storage size (in HSM)

Analysis time TAG	Link speed to HSM
Analysis time AOD	AMS write speed
Analysis time ESD	AMS read speed
Memory for RAW->ESD processing job	(maximum disk read/write speed)
Memory for ESD->AOD processing job	
Memory for AOD->TAG processing job	**data access pattern parameters (local)**
Memory for TAG analysis job	Fraction of events for which TAG->AOD associations are followed
Memory for AOD analysis job	
Memory for ESD analysis job	Fraction of events for which AOD->ESD associations are followed
Container size RAW	
Container size ESD	Fraction of events for which ESD->RAW associations are followed
Container size AOD	
Container size TAG	Clustering density parameter

A number of parameters can be modified easily using the GUI menus, they include most of the **global** parameters describing the analysis (CPU needed by various JOBS, as well as memory required for processing) and most of **local** parameters defining the hardware and network configuration of each of the regional centres which are part of the model (an arbitrary number of regional centres can be simulated, each with different configuration and with different data residing on it). Also, the basic hardware costs can be input via GUI, which allows simple estimates of the overall cost of a system. This part of the simulation program will certainly evolve to include the price for items which are more difficult to quantify, like inconvenience and discomfort, travel costs et cetera. For each regional centre, one can define a different set of jobs to be run. In particular, one could define different data access patterns in physics analyses performed in each of the centres, with different frequencies of following the TAG->AOD, AOD->ESD and ESD->RAW associations.

Validation of MONARC simulation tool

A set of measurements of the key parameters of the distributed database, such as AMS read/write speeds with a single user and also with stress tests have been measured. A close discussion between the Analysis WG and Testbed WG helped to identify the key parameters and the dependencies of the parameters needed in the simulation program. The correctness of the scaling behaviour, in both the local and wide-area network environment, which is vital in making any predictions on a large scale distributed system, has been validated successfully in Summer 1999. Later, a series of simulation results were compared against analytical calculations based on queueing theory. Perfect agreement was found.

Milestones

The MONARC project has successfully met all its major milestones, and is well on the way to meeting its primary goals, including

- Identifying first-round baseline Computing Models that could provide viable (and cost-effective) solutions to meet the data analysis needs of the LHC experiments
- Providing a powerful (CPU and time efficient) simulation tools that will enable further computing model studies
- Providing guidelines for the configuration and services of Regional Centers, and
- Providing an effective forum where representatives of actual and candidate Regional Centers may meet and develop common strategies for LHC computing.

A series of short papers on: MONARC simulation tool; the structure and operational experience with the simulation system (using the results of the Analysis Working Group); the work of the Architecture Working Group; and the testbed studies and simulation validaton in local and wide-area network environments will be presented at CHEP2000 Conference.

Deliverables

All existing information, including various presentations in which the logical model of the MONARC simulation tool has been has presented, some documentation, simple examples and demos are available on from MONARC WWW pages (MONARC->Simulation and Modelling->Status of the Simulation Software). All results from various simulations runs, together with all parameter sets as well as Java classes used in those runs, are available from the WWW pages. A detailed user manual to MONARC simulation is under development, and should be available in the Spring 2000.

MONARC Phase 3

It is believed that from 2000 onwards, a significant amount of work will be necessary to model, prototype and optimise the design of the overall distributed computing and data handling systems for the LHC experiments. This work, much of which should be done in common for the experiments, would be aimed at providing "cost effective" means of doing data analysis in the various world regions, as well as at CERN. Finding common solutions would save some of the resources devoted to determining the solutions, and would ensure that the solutions found were mutually compatible. The importance of compatibility based on common solutions applies as much to cases where multiple Regional Centres in a country intercommunicate across a common network infrastructure, as it does to sites (including CERN) that serve more than one LHC experiment.

A letter of intent to continue with MONARC Phase 3 has been sent to H. Hoffman and M. Delfino in December 1999. MONARC Phase 3 could have a useful impact in several areas, including:

- facilitation of contacts, discussions, interchanges, for the planning and mutually compatible design of centre and network architecture and services (among the experiments, the CERN Centre and the Regional Centres)
- modelling consultancy and "service" to the experiments and Centres
- providing a core of advanced R&D activities encompassing system optimisation, and pre-production prototyping
- taking advantage of the work on distributed data-intensive computing systems beginning this year in other "next generation" R&D projects

Details on the synergy between a MONARC Phase 3 and R&D projects such as the recently approved Next Generation Internet "Particle Physics Data Grid" (PPDG) may be found in [29]. The PPDG project (involving ANL, BNL, Caltech, FNAL, JLAB, LBNL, SDSC, SLAC, and the University of Wisconsin) shares MONARC's aim of finding common solutions to meet the large-scale data management needs of high energy (as well as nuclear) physics. Some of the concepts of a possible Phase 3 study are briefly summarised below.

The Phase 3 study could be aimed at maximising the workload sustainable by a given set of networks and site facilities, or at reducing the long turnaround times for certain data analysis tasks, or a combination of both. Unlike Phase 2, the optimization of the system in Phase 3 would no longer exclude long and involved decision processes, as the potential gains in terms of work accomplished or resources saved could be large. Some examples of the complex elements of the Computing Model that might determine the (realistic) behaviour of the overall system, and which could be studied in Phase 3 are

- **Resilience**, resulting from flexible management of each data transaction, especially over wide area networks
- **Fault tolerance**, resulting from robust fall-back strategies and procedures (automatic and manual, if necessary) to recover from abnormal conditions (such as irrecoverable error conditions due to data corruption, system thrashing, or a subsystem falling offline).
- **System state tracking**, so that the capability of the system to respond to requests is known (approximately) at any given time, and the time to satisfy requests for data and/or processing power may be, on average, reliably estimated, or abnormal conditions may be detected and in some cases predicted.

MONARC in Phase 3 could exploit the studies, system software developments, and prototype system tests completed by early 2000, to develop more sophisticated and efficient Models than were possible in Phase 2. The Simulation and Modelling work of MONARC on data-intensive distributed systems is likely to be more advanced than in PPDG or other NGI projects in 2000, so that MONARC Phase 3 could have a central role in the further study and advancement of the

design of distributed systems capable of PetaByte-scale data processing and analysis. As mentioned in the PEP, this activity would potentially be of great importance not only for the LHC experiments, but for scientific research on a broader front, and eventually for industry.
Tufts University Group will continue its participation in MONARC Phase 3.

ATLAS-specific and US_ATLAS specific simulations with MONARC toolset

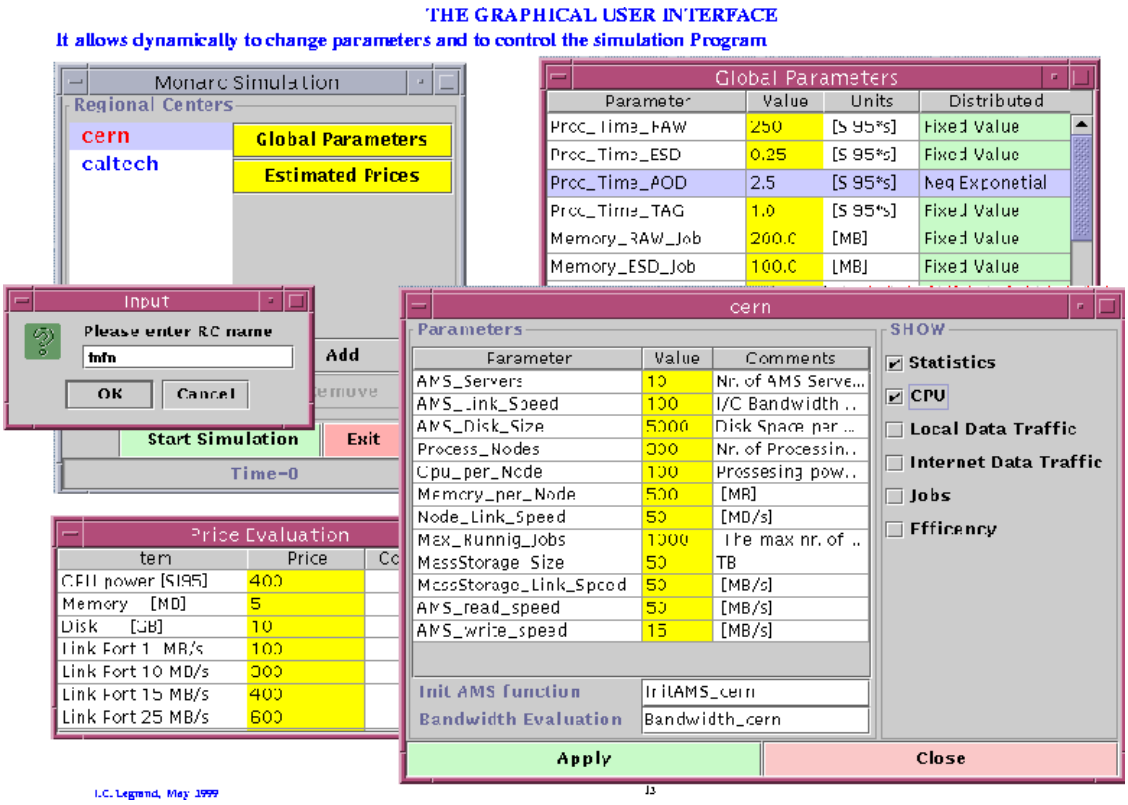
Tufts University group will lead both tasks. ATLAS Computing Co-ordinator, N. McCubbin, has asked K.Sliwa to lead a still to-be-formed group which will conduct a systematic study and optimization of possible architectures of ATLAS computing system. Tufts group has all the needed expertise and experience with both the MONARC simulation tools and the problem of optimizing data access to facilitate physics analyses. Analogous optimization studies will be performed for US-ATLAS needs, in which case emphasis should be put on optimizing the balance between Tier-1 and Tier-2 centers. Continuing support for a computing expert/analyst at Tufts University is requested. It is anticipated that results of the first-round studies will be available in the end of 2000.

Appendix B: References

1) The WWW Home Page for the MONARC Project
http://www.cern.ch/MONARC/
2) The MONARC Project Execution Plan, September 1998
http://www.cern.ch/MONARC/docs/pep.html
3) Christoph von Praun: Modelling and Simulation of Wide Area Data Communications.
A talk given at the CMS Computing Steering Board on 19/06/98.
4) J.J.Bunn: Simple Simulation of the Computing Models:
http://pcbunn.cithep.caltech.edu/results/model/model.html
5) The PTOLEMY Simulation Tool:
http://www-tkn.ee.tu-berlin.de/equipment/sim/ptolemy.html
6) The SES Workbench
http://www.ses.com/Workbench/index.htm
7) **PARASOL** - C/C++ simulation library for dist / parallel systems
http://www.hensa.ac.uk/parallel/simulation/architectures/parasol/index.html
8) Rough Sizing Estimates for a Computing Facility for a Large LHC experiment, Les Robertson. MONARC-99/1.
http://nicewww.cern.ch/~les/monarc/capacity_summary.html.
9) Report on Computing Architectures of Existing Experiments, V.O'Dell et al. MONARC-99/2.
http://home.fnal.gov/~odell/monarc_report.html
10) Regional Centers for LHC Computing, Luciano Barone et al. MONARC-99/3. (text version)
http://home.cern.ch/~barone/monarc/RCArchitecture.html
11) Presentations and notes from the MONARC meeting with Regional Center Representatives April 23, 1999
http://www.fnal.gov/projects/monarc/task2/rc_mtg_apr_23_99.html
12) PASTA, Technology Tracking Team for Processors, Memory, Storage and Architectures:
http://nicewww.cern.ch/~les/pasta/run2/welcome.html
13) Home page of the Analysis Design Working Group:
http://www.bo.infn.it/monarc/ADWG/AD-WG-Webpage.html
14) Analysis Processes of current and imminent experiments:
http://www.bo.infn.it/monarc/ADWG/Meetings/Docu-15-12-98.html
15) Monarc Note 98/1:
http://www.mi.infn.it/~cmp/rd55/rd55-1-98.html
16) CMS TN-1996/071 The CMS Computing Model
17) Parameters of the initial Analysis Model:
http://www.bo.infn.it/monarc/ADWG/Meetings/15-01-99-Docu/Monarc-AD-WG-0199.html

18) Unfeasable models evaluations:
http://www.bo.infn.it/monarc/ADWG/Meetings/Docu-24-01-99.html (to be released)
19) Preliminary evaluation criteria (slide 8)
http://bo_srv1_nice.bo.infn.it/~capiluppi/monarc-workshop-0599.pdf
20) ATLFAST++ in LHC++:
http://www.cern.ch/Atlas/GROUPS/SOFTWARE/OO/domains/analysis/atlfast++.html
21) GIOD (Globally Interconnected Object Databases) project:
http://pcbunn.cithep.caltech.edu/Default.htm
22) ATLAS 1 TB Milestone:
http://home.cern.ch/s/schaffer/www/slides/db-meeting-170399-new/
23) CMS TestBeam web Page
http://cmsdoc.cern.ch/ftp/afscms/OO/Testbeams/www/Welcome.html
24) MONARC-99/4: M. Boschini, L. Perini, F. Prelz, S. Resconi: Preliminary Objectivity tests for MONARC project on a local federated database:
http://www.cern.ch/MONARC/docs/monarc_docs/99-04.ps
25) K. Holtman: CPU requirements for 100 MB/s writing with Objectivity:
http://home.cern.ch/~kholtman/monarc/cpureqs.html
26) K. Amako, Y. Karita, Y. Morita, T. Sasaki and H. Sato: MONARC testbed and a preliminary measurement on Objectivity AMS server:
http://www-ccint.kek.jp/People/morita/Monarc/amstest.ps
27) K. Sliwa: What measurements are needed now?:
http://www.cern.ch/MONARC/simulation/measurements_may_99.htm
28) C. Vistoli: QoS Tests and relationship with MONARC:
http://www.cnaf.infn.it/~vistoli/monarc/index.htm
29) H. Newman: Ideas for Collaborative work as a Phase 3 of MONARC
http://www.cern.ch/MONARC/docs/progress_report/longc7.html

Figure 1. An example of GUI-based setup of a MONARC simulation run.



THE GRAPHICAL USER INTERFACE

It allows dynamically to change parameters and to control the simulation Program